Ecology, 102(2), 2021, e03241

Generalized model-based solutions to false-positive error in species detection/nondetection data

John D. J. Clare [1], Philip A. Townsend, and Benjamin Zuckerberg

Department of Forest and Wildlife Ecology, University of Wisconsin—Madison, 1630 Linden Drive, Madison, Wisconsin 53706 USA

Citation: Clare, J. D. J., P. A. Townsend, and B. Zuckerberg. 2021. Generalized model-based solutions to false-positive error in species detection/nondetection data. Ecology 102(2):e03241. 10.1002/ecy.3241

Abstract. Detection/nondetection data are widely collected by ecologists interested in estimating species distributions, abundances, and phenology, and are often imperfect. Recent model development has focused on accounting for both false-positive and false-negative errors given evidence that misclassification is common across many sampling protocols. To date, however, model-based solutions to false-positive error have largely addressed occupancy estimation. We describe a generalized model structure that allows investigators to account for false-positive error in detection/nondetection data across a broad range of ecological parameters and model classes, and demonstrate that previously developed model-based solutions are special cases of the generalized model. Simulation results demonstrate that estimators for abundance and migratory arrival time ignoring false-positive error exhibit severe (20–70%) relative bias even when only 5–10% of detections are false positives. Bias increased when false-positive detections were more likely to occur at sites or within occasions in which true positive detections were unlikely to occur. Models accounting for false-positive error following the site-confirmation or observation-confirmation designs generally reduced bias substantially, even when few detections were confirmed as true or false positives or when the process model for false-positive error was misspecified. Results from an empirical example focusing on gray fox (*Urocyon cinereoargenteus*) abundance in Wisconsin, USA reinforce concerns that biases induced by false-positive error can also distort spatial predictions often used to guide decision making. Model sensitivity to false-positive error extends well beyond occupancy estimation, but encouragingly, model-based solutions developed for occupancy estimators are generalizable and effective across a range of models widely used in ecological research.

Key words: abundance; imperfect detection; misclassification; monitoring; phenology; precision/recall; species distribution model.

Introduction

Detection/nondetection data are widely collected by ecologists interested in monitoring populations or elucidating habitat associations (MacKenzie et al. 2002). It is now widely recognized that species detection/nondetection or occurrence data and many types of ecological survey data suffer from imperfect detection: The actual occurrence of species or individuals is rarely perfectly observed. Concerns about false-negative error, the failure to detect an individual or species when present, have been recognized for decades and have motivated the development of a broad range of models that account for it explicitly (MacKenzie et al. 2002, Royle and Nichols 2003).

False-positive error, another type of imperfect detection within species occurrence data, can occur when a nontarget species or another phenomena (e.g., extrinsic sound) is misclassified as the focal species of interest (McClintock et al. 2010). Molecular assays for infectious agents face similar issues, as sample contamination or nonspecific amplification can result in false-positive test results (Brost et al. 2018). Across a variety of species and sampling protocols, the incidence of false-positive error has been estimated as varying from nearly negligible to constituting 20% of observations or more (McClintock et al. 2010, Swanson et al. 2016). Simulation results have shown that even relatively few false-positive detections can severely bias estimates of species occupancy when models account for false negatives but ignore false positives (Miller et al. 2011, Ruiz-Gutiérrez et al. 2016).

The prevalence of false-positive error has spurred investigators to adopt a variety of strategies aimed at ameliorating potential biases. For species detection/nondetection data, strategies include a complete data review after collection, data collection or processing methods aimed at reducing the incidence of false positives, and model-based approaches. Implementing complete data reviews requires all data to be reviewable and can be burdensome or infeasible for large data sets (Gardiner et al.

Manuscript received 18 December 2019; revised 18 August 2020; accepted 14 September 2020. Corresponding Editor: Viviana Ruiz-Gutierrez.
[1] E-mail: john.d.j.clare@gmail.com

2012, Ruiz-Gutiérrez et al. 2016). Specific data collection or processing protocols aimed at reducing the incidence of false positives include performing partial data reviews to develop indicators or algorithms for identifying false positives, simplifying classification tasks, or providing additional guidance or training to human or computer-based classifiers (Miller et al. 2012*a*, Kosmala et al. 2016, Swanson et al. 2016). Although these approaches can greatly improve data quality, they exhibit certain inefficiencies. It can be time consuming to quantify how accurately data have been classified, what level of classification accuracy is sufficient for a specific research objective, and whether manipulating training or classification protocols result in sufficient classification accuracy.

Model-based solutions are a more efficient way to ameliorate false-positive error, and several variants for detection/nondetection data have been described specifically for occupancy models (Royle and Link 2006, Miller et al. 2011, Chambert et al. 2015, Ferguson et al. 2015, Ruiz-Gutiérrez et al. 2016, Brost et al. 2018). The original occupancy model described by MacKenzie et al. (2002) conceptualizes observed absences as a mixture of true and false negatives. Occupancy models that account for false-positive error conceptualize detections as a mixture of true and false positives. Under the "full" estimator described by Royle and Link (2006), all observed occurrences are of unknown reliability, and disentangling the false-negative and false-positive mixtures requires constrained priors. Subsequent developments leverage auxiliary data collected under different protocols to improve discrimination between true and false positives. Investigators following the site-confirmation design are able to unambiguously classify some detections as true positives (e.g., via a posteriori confirmation or by paired sampling with using a method free from false-positive error), whereas other detections are ambiguously true positives or false positives (Miller et al. 2011, Ferguson et al. 2015). The observation-confirmation design is an extension of the site-confirmation protocol in which some detections can be unambiguously classified as true positives or false positives (e.g., via a posteriori laboratory tests; Chambert et al. 2015). The calibration design involves assessing classification performance within settings in which the ecological state variable is experimentally controlled (e.g., playback experiments or negative laboratory controls; Ruiz-Gutiérrez et al. 2016, Brost et al. 2018). These model-based solutions alleviate bias in occupancy estimators while sparing time associated with performing complete data reviews, making changes to classifier training, or developing and calibrating error indicators.

The development and application of model-based solutions to false-positive error has largely focused on occupancy estimation. However, species detection/nondetection data are increasingly used to estimate state variables other than occupancy, including abundance, phenology, and associated dynamics (Royle and Nichols 2003, Chandler and Clark 2014, Roth et al. 2014, Ramsey et al. 2015, Rossman et al. 2016). Presumably, these estimators are similarly biased by false-positive error, and these biases could severely hamper a broad set of ecological decisions ranging from assessing the recovery of protected populations to delineating seasonal protections for migratory species.

Here, we show that previously developed model-based solutions to false-positive error can be described as specific cases of a generalized model to account for false-positive error in detection/nondetection data. We use simulation to quantify the bias caused by false-positive detections across estimators of different state variables like abundance or phenological phenomena such as migratory arrival or emergence from hibernation, and to demonstrate that model-based solutions commonly improve inference across a range of estimation problems. Although generalization is not restricted to any specific design, we primarily focus on study designs employing observation confirmation (Chambert et al. 2015), which is the most applicable when researchers use sampling techniques that produce data that can be reviewed and verified a posteriori.

Generalizing Model-Based Solutions for False Positives

Let $\mathbf{y}$ denote a matrix of binary observations corresponding to the detection or nondetection of a species at $i = 1, 2, \ldots, R$ locations over $j = 1, 2, \ldots, T$ discrete sampling occasions. If a species or more specific species state of interest such as a juvenile is observed, $y_{i,j} = 1$, and $y_{i,j} = 0$ otherwise. We proceed assuming that $\mathbf{y}$ is *repeated* detection/nondetection data. However, the concepts apply across data with different dimensions or for data collected slightly differently. For example, $\mathbf{y}$ could denote a vector of presence–background data, where $y_i = 1$ indicates that the species was detected or occurred at location i, and where $y_i = 0$ indicates that location i is part of the randomly selected background sample.

Parametric models for detection/nondetection data typically assume $y_{i,j} \sim \text{Bernoulli}(\theta)$, where θ is the unconditional probability of detection at a specific place and time. This probability is equivalent to the union of the respective unconditional probabilities of true-positive detection (θ_{tp}) and false-positive detection (θ_{fp}). If $\theta_{fp} = 0$, $\theta \equiv \theta_{tp}$ and the union between true- and false-positive detections does not need to be explicitly specified. For example, MacKenzie et al. (2002) define the unconditional probability of detection $\theta_i = z_i \times p$, where z_i is the latent binary occupancy state of site i distributed as Bernoulli (ψ), ψ is a probability of occupancy that might vary in relation to site-level covariates, and p is the probability of detecting an organism at a site given that it is present and might vary in relation to either site and occasion level covariates. An equivalent but less compact description is that $\theta_{fp} = 0$, $\theta_{tp,i} = z_i \times p$, and $y_{i,}$

$_j \sim$ Bernoulli ($\theta_{tp,i} \cup \theta_{fp}$). If $\theta_{fp} > 0$, decomposing θ into true- and false-positive probabilities is critical for unbiased estimation of θ_{tp}, which is typically the focus of ecological inquiry. A model-based solution to false-positive error within species detection/nondetection data is simply any model that explicitly assumes

$$y_{i,j} \sim \text{Bernoulli}(\theta_{tp} \cup \theta_{fp}) \quad (1)$$

and estimates these probabilities or their constituent parameters. Specific solutions may differ with respect with respect to how θ_{tp}, $\theta_{fp,}$ and their union are defined, and other estimation details.

The union between θ_{tp} and θ_{fp} may take a few forms. The first is what we refer to as an inclusive union or form. In certain sampling situations, false positives might be possible at both the site level (detections at unoccupied sites) and the observation level (false detections at occupied sites). Further, it might be possible for a target species or entity to be truly and falsely detected at the same location during the same sampling occasion. In particular, if a spatial or temporal interval is used to define a occasion (e.g., 1 km of transect, a 24-h interval), the target species may be detected correctly or falsely multiple times. When collapsing this count of detections into binary data denoting >0 detections or not, any $y_{i,}$ $_j = 1$ may include true and/or false positives. Assuming true and false positives occur independently or are conditionally independent given a set of covariates, $\theta_{tp} \cup \theta_{fp} = \theta_{tp} + \theta_{fp} - (\theta_{tp} \times \theta_{fp})$. In turn, the distribution of the collapsed detection/nondetection data can be described as

$$y_{i,j} \sim \text{Bernoulli}(\theta_{tp} + \theta_{fp} - [\theta_{tp} \times \theta_{fp}]). \quad (2)$$

The union above is a factorization of three distinct probabilities (Chambert et al. 2015, Brost et al. 2018): the probability of >0 true positives only ($\theta_{tp} \times [1 - \theta_{fp}]$), the probability of >0 false positives only ($\theta_{fp} \times [1 - \theta_{tp}]$), and the probability of >0 true positives and >0 false positives ($\theta_{tp} \times \theta_{fp}$). For example, at a single camera trap during a day-long sampling occasion, the target species might be recorded and correctly identified, a nontarget species might be recorded and misidentified as the target, or both. An occupancy model accounting for inclusive false-positive error is a particular case of (1) in which $\theta_{tp,i} \cup \theta_{fp} = \theta_{tp,i} + \theta_{fp} - (\theta_{tp,i} \times \theta_{fp})$, $\theta_{tp,i} = z_i \times p$, and where θ_{fp} is a constant or some combination of parameters to be estimated.

In other situations where site-level and observation-level false positives are possible, true and false positives might be mutually exclusive events, in which a detection can only be a true positive or a false positive. Here, the product $\theta_{tp} \times \theta_{fp} = 0$, so $\theta_{tp} \cup \theta_{fp} = \theta_{tp} + \theta_{fp}$ and

$$y_{i,j} \sim \text{Bernoulli}(\theta_{tp} + \theta_{fp}). \quad (3)$$

We refer to this as an observationally exclusive union or form. For example, a camera trap might be programmed to take a single time-lapse image per day-long sampling occasion, in which case the target species might be detected and correctly classified, or some other species might be detected and misclassified as the target species, but not both (assuming one organism within the image). An occupancy model accounting for observationally exclusive false-positive error is a particular case of (1) in which $\theta_{tp,i} \cup \theta_{fp} = \theta_{tp,i} + \theta_{fp}$, $\theta_{tp,i} = z_i \times p$ and θ_{fp} is a constant or combination of parameters to be estimated. Note that one can use (3) rather than (2) even if false-positive error is inclusive, provided one interprets a false positive as any $y_{i,j} = 1$ where zero observations are true positive. In this case, θ_{fp} represents the probability that *all* observations at some location and occasion are false positives. Interpretation and estimation of θ_{tp} is unchanged.

A specific form that is most commonly used within model-based solutions for false-positive error makes assumptions we refer to as "conditionally exclusive." As with the observationally exclusive union, $y_{i,j} \sim$ Bernoulli ($\theta_{tp} + \theta_{fp}$). What differentiates this form is that θ_{tp} and θ_{fp} are not only mutually exclusive within any cell $y_{i,j}$, but that there are spatial or temporal conditions under which only true positive observations are possible, and if these conditions are not met, all observations are false positives. This formulation is commonly employed within occupancy models accounting for false-positive error following the full, site confirmation, or calibration designs. Probabilities p_{tp} and p_{fp} (often referred to as p_{11} and p_{10}) are conditional upon whether site i is occupied ($z_i = 1$) or not ($z_i = 0$), respectively (Miller et al. 2011). These are particular cases of (1) in which $\theta_{tp,i} \cup \theta_{fp,i} = \theta_{tp,i} + \theta_{fp,i}$, $\theta_{tp,i} = z_i \times p_{tp}$, and $\theta_{fp,i} = (1 - z_i) \times p_{fp}$. Although the assumption that only site-level false positives can occur is probably violated in many sampling situations, occupancy estimation is not strongly biased by such violations (although coverage suffers; Ferguson et al. 2015, Brost et al. 2018).

Estimating Θ_{TP} and Θ_{FP} Following the Observation-Confirmation Design

The observation-confirmation design requires investigators to review and confirm all detections within some subset of occasions at some subset of sites a posteriori. Confirmed data $\mathbf{v}$ are classified such that $v_{i,j} = 1$ if all detections at a site and occasion were confirmed to be true positives, $v_{i,j} = 2$ if all were confirmed as false positives, and $v_{i,j} = 3$ if confirmation reveals both true- and false-positive detections. If the species is not detected at all and there is nothing to verify, $v_{i,j} = 0$.

Chambert et al. (2015) assume $v_{i,j} \sim$ Categorical ($\mathbf{\Omega_i}$) and condition vector $\mathbf{\Omega_i}$ upon whether the species occurs at site i or not. Let s_{tp} and s_{fp} denote the respective

probabilities that a sampling interval contains >0 true positive detections at an occupied site or >0 false positives. If $z_i = 1$ (species occurs), it may not be detected, it can be truly detected only, falsely detected only, or detected both ways, and $\mathbf{\Omega_i} = [\{(1 - s_{tp}) \times (1 - s_{fp})\}\ \{(s_{tp} \times (1 - s_{fp})\}\ \{(1 - s_{tp}) \times s_{fp}\}\ \{s_{tp} \times s_{fp}\}]$; brackets {} are used to delineate the distinct scalars in $\mathbf{\Omega_i}$. If $z_i = 0$, the only possible outcomes are no detection or a false-positive detection. Removing conditioning upon a specific occupancy state and substituting $\theta_{tp,i}$ for $z_i \times s_{tp}$ and θ_{fp} for s_{fp} yields that $\mathbf{\Omega_i} = [\{(1 - \theta_{tp,i}) \times (1 - \theta_{fp})\}\ \{\theta_{tp,i} \times (1 - \theta_{fp})\}\ \{(1 - \theta_{tp,i}) \times \theta_{fp}\}\ \{\theta_{tp,i} \times \theta_{fp}\}]$. The first scalar is equal to the probability of nondetection, and the latter three are the probabilities of specific detection outcomes factorized in (2).

In turn, Chambert et al. (2015) assume unconfirmed detection/nondetection data $y_{i,j} \sim \text{Bernoulli}(z_i \times [s_{tp} + s_{fp} - (s_{tp} \times s_{fp})] + [1 - z_i] \times s_{fp})$. If present ($z_i = 1$), a species can be detected ($y_{i,j} = 1|z_i = 1$) either truly, falsely, or both: $s_{tp} + s_{fp} - s_{tp} \times s_{fp}$ denotes the factorization of these conditional probabilities. If the species is not present ($z_i = 0$), it can only be falsely detected ($y_{i,j} = 1|z_i = 0$). Substituting $\theta_{tp,i,}$ for $z_i \times s_{tp}$ and θ_{fp} for s_{fp} yields that $y_{i,j} \sim \text{Bernoulli}\ (\theta_{tp,i} + \theta_{fp} - [\theta_{tp,i} \times \theta_{fp}])$. This is exactly Eq. 2.

That is, an occupancy model following the observation-confirmation protocol is a special case of Eq. 2 where $\theta_{tp,i} = z_i \times s_{tp}$ and $\theta_{fp} = s_{fp}$. Note that s_{tp} is equivalent to what MacKenzie et al. (2002) call p. A hierarchical description of the complete data likelihood is then

$$
\begin{aligned}
z_i &\sim \text{Bernoulli}(\psi)\\
\theta_{tp,i} &= z_i \times p\\
\Omega_i &= [\{(1-\theta_{tp,i}) \times (1-\theta_{fp})\}\{\theta_{tp,i} \times (1-\theta_{fp})\}\\
&\qquad \{(1-\theta_{tp,i}) \times \theta_{fp}\}\{\theta_{tp,i} \times \theta_{fp}\}]\\
v_{i,j} &\sim \text{Categorical}(\Omega_i)\\
y_{i,j} &\sim \text{Bernoulli}(1-\Omega_{1,i})
\end{aligned}
$$

A point which we will return to is that $\theta_{tp,i}$ is the same as the unconditional probability of detection—$\Pr(y_{i,j} = 1)$—presented by MacKenzie et al. (2002).

First we acknowledge some superficial differences between the presentation here and by Chambert et al. (2015; Table 1). They denote s_{tp} and s_{fp} as s_1 and s_0, respectively. They present p_{10} as the probability $y_{i,j} = 1|z_i = 0$, but is redundant ($p_{10} = s_0$ [or s_{fp}] $= \theta_{fp}$) and only one term is needed. Similarly, s_{tp} (or s_1) is redundant with the original model's p (MacKenzie et al. 2002). Chambert et al. (2015) derive $p_{11} = s_{tp} + s_{fp} - (s_{tp} \times s_{fp})$ to describe $\Pr(y_{i,j} = 1|z_i = 1)$, but this derivation is not strictly necessary. Importantly, notation s_0, p_{10}, and p_{11} also lacks a consistent interpretation within the existing literature (Table 1). The site-confirmation and calibration protocols typically condition s_0 and p_{10} on absence, and p_{11} strictly represents a true positive probability conditional upon presence. The observation-confirmation protocol defines s_0 as an unconditional probability of falsely detecting a species and defines p_{11} as the probability of detection (either true or false positive) at occupied sites (Chambert et al. 2015).

We assume inclusive false-positive error in the extensions described below, but an observationally exclusive formulation could be implemented by assuming $\theta_{tp,i} \times \theta_{fp} = 0$, and reducing $\mathbf{\Omega_i}$ to describe three detection states corresponding to no detection, a true positive only, or a false positive only (Appendix S1, Fig. S1): $\mathbf{\Omega_i} = [\{1 - (\theta_{tp,i} + \theta_{fp})\}\ \{\theta_{tp,i}\}\ \{\theta_{fp}\}]$.

Table 1. Variables and symbology used within the original descriptions of the observation-confirmation occupancy model and our reformulation here, and where applicable, differences between the observation-confirmation protocol and other protocols.

Symbol	Description
z_i	Latent binary occupancy state of site i
$y_{i,j}$	Binary detection/nondetection at site i, occasion j; detections not verified and potentially include false positives
$v_{i,j}$	Categorical data derived from post hoc verification of all observations at site i, occasion j: true-positive detection(s) only, false-positive detection(s) only, true- and false-positive detection(s), no detection
p_{10}	A parameter describing the probability that $y_{ij} = 1\|\ z_i = 0$ (i.e., detection at unoccupied sites). Because detections at unoccupied sites can only be false positive, $p_{10} = s_0$
p_{11}	A parameter describing the probability that $y_{ij} = 1\|\ z_i = 1$ (i.e., detection at occupied sites). When following the observation conformation protocol, this includes true and false positives ($p_{11} = s_{tp} + s_{fp} - [s_{tp} \times s_{fp}]$). When following the site-confirmation or calibration protocols, p_{11} is equal to s_{tp} (or the parameter p used in a standard occupancy model), because all detections at occupied sites are assumed to be true positive detections
s_{tp} (s_1)	Following observation-confirmation and calibration protocols, a parameter that describes the probability of a true-positive detection (i.e., >0 confirmed true positive observations within v_{ij}) given that a site is occupied. Equivalent to the parameter p used in a standard occupancy model
s_{fp} (s_0)	Following observation-confirmation protocol, a parameter describing the unconditional probability of a false-positive detection (i.e., >0 confirmed false-positive observations within v_{ij}). Under calibration protocol, may be conditional upon site being unoccupied (Chambert et al. 2015). Brost et al. (2018) describe the parameter unconditionally and as equivalent to θ_{fp}, but call it φ
θ_{tp}	Following description here, a derived parameter describing the unconditional probability of a true-positive detection within both y_{ij} and v_{ij}. $\theta_{tp,i} = z_i \times s_1$ (or $z_i \times p$)
θ_{fp}	Following description here, a parameter describing the unconditional probability of a false-positive detection within both $y_{i,j}$ and $v_{i,j}$

Extension to Nonoccupancy Models

If assuming an inclusive or observationally exclusive union between true and false positives and following the observation-confirmation design, the only difference between the occupancy model described above and a different model accounting for false-positive error relates to how $\boldsymbol{\theta}_{\mathbf{tp}}$ is defined (Fig. 1; Appendix S1, Fig. S1). Extension to other model classes requires deriving $\boldsymbol{\theta}_{\mathbf{tp}}$ as the unconditional probability of detection presented in the original model description, and appropriately specifying $\boldsymbol{\theta}_{\mathbf{tp}}$'s constituent processes. We present two examples below.

Royle–Nichols model

Royle and Nichols (2003; RN model hereafter) describe the unconditional probability of detection (i.e., $\theta_{tp,i}$) as $1-(1-r)^{N_i}$, where r is the probability of detecting an individual during a sampling interval, and N_i, distributed as Poisson (λ), denotes the abundance of a species at site i. The hierarchical likelihood for a version

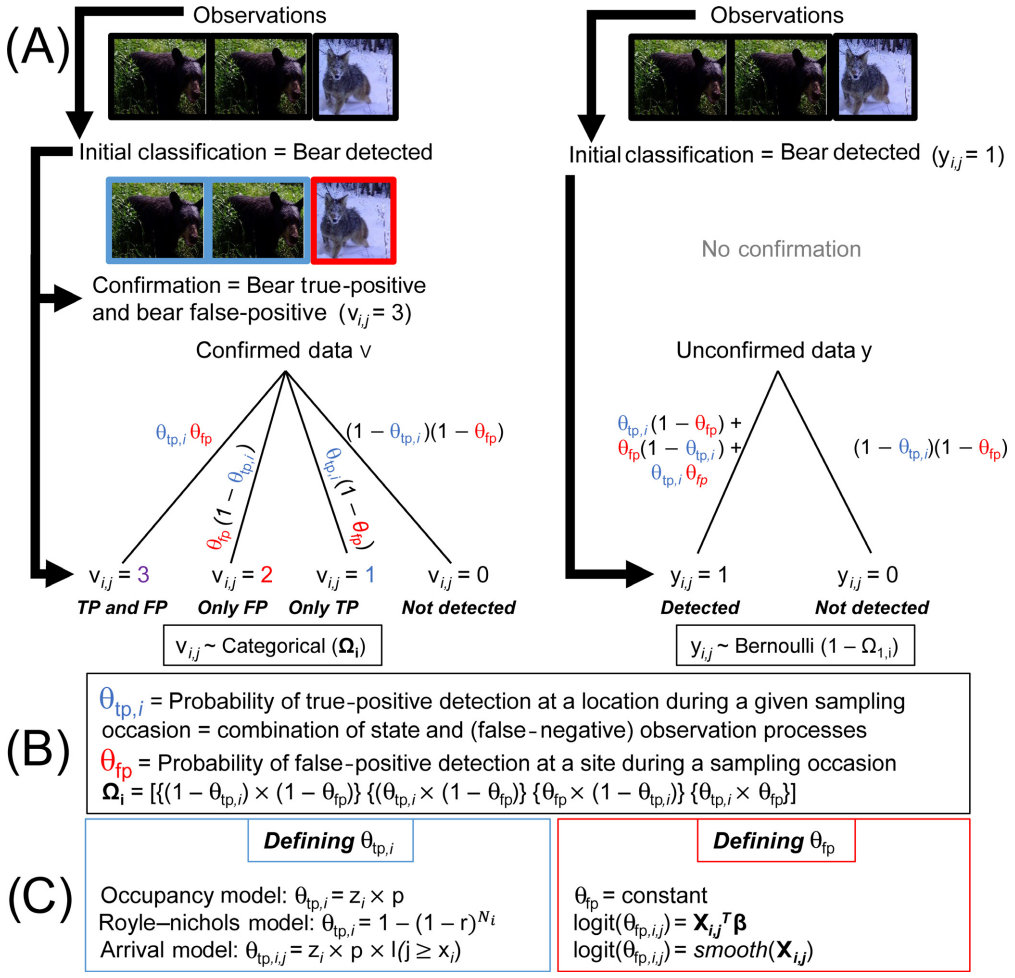

Fig. 1. Schematic outlining how the observation-confirmation design can be implemented to deal with false-positive error across several model classes. The design presupposes that multiple observations at any site and occasion—which may include both true (TP) and false positives (FP) following Eq. 2—are collapsed into binary detection/nondetection (A). At left, detections at a subset of sites and sampling occasions are confirmed a posteriori: the example here depicts that at a specific site and sampling occasion, both true (two bear images) and false (one coyote image) positives occur, and the confirmed observations are collapsed into categorical data: here, $v_{i,j} = 3$. Unconfirmed data $y_{i,j}$ (example at right) is either classified as 0 (no detection) or 1 (detected). The probability that $y_{i,j} = 1$ is equivalent to the sum of the probabilities that the observations constituting a detection at site i and occasion j consist of entirely true positives ($v_{i,j} = 1$), entirely false positives ($v_{i,j} = 2$), or a mix of both ($v_{i,j} = 3$). The probabilities underpinning y and v reflect mixtures of the unconditional probabilities of true- (θ_{tp}) and false-positive (θ_{fp}) detection (B). The unconditional probability of a true-positive detection is the same as the unconditional probability of detection—that is, $\Pr(y_{i,j} = 1)$—defined within the base model of interest (C). In turn, because true and false positives are assumed to be independent, the unconditional probability of a false-positive detection can be modeled as a constant or as functionally varying in relation to site, occasion, or site-by-occasion covariates irrespective of the model for θ_{tp}. See Appendix S1: Fig. S1 for an analogous figure following Eq. 3.

assuming inclusive false-positive error following the observation-confirmation protocol is then

$$
\begin{aligned}
&N_i \sim \text{Poisson}(\lambda)\\
&\theta_{\text{tp},i} = 1-(1-r)^{(N_i)}\\
&\Omega_i = [\{(1-\theta_{\text{tp},i})\times(1-\theta_{\text{fp}})\}\{\theta_{\text{tp},i}\times(1-\theta_{\text{fp}})\}\\
&\qquad \{(1-\theta_{\text{tp},i})\times\theta_{\text{fp}}\}\{\theta_{\text{tp},i}\times\theta_{\text{fp}}\}]\\
&v_{i,j} \sim \text{Categorical}(\Omega_i)\\
&y_{i,j} \sim \text{Bernoulli}(1-\Omega_{1,i})
\end{aligned}
$$

The only differences between this model and the occupancy model presented in the previous section are within the first two lines: a process model for N_i replaces a process model for z_i, and $\theta_{\text{tp},i}$ is derived as a function of r and N_i rather than z_i and p.

Phenological "arrival" model

Incorporating false positives within an occupancy model designed to estimate the timing of some ephemeral phenomena such as migration arrival or emergence from torpor (Roth et al. 2014; PA model hereafter) follows the Chambert et al. (2015) description, except that organisms can only be truly detected during sampling occasions at occupied sites after arrival. Thus, $\boldsymbol{\theta}_{\mathbf{tp}}$ must be described using indexing for i locations and j time periods (i.e., $\theta_{\text{tp},i,j}$). Let arrival time at site i be denoted as x_i and assume that $x_i \sim$ Poisson (φ). To simplify presentation, we define x_i in terms of sampling intervals j rather than specific dates. The hierarchical likelihood is

$$
\begin{aligned}
&z_i \sim \text{Bernoulli}(\psi)\\
&x_i \sim \text{Poisson}(\varphi)\\
&\theta_{\text{tp},i,j} = z_i \times p \times I(j \geq x_i)\\
&\Omega_{\mathbf{i,j}} = [\{(1-\theta_{\text{tp},i,j})\times(1-\theta_{\text{fp}})\}\{\theta_{\text{tp},i,j}\times(1-\theta_{\text{fp}})\}\\
&\qquad \{(1-\theta_{\text{tp},i,j})\times\theta_{\text{fp}}\}\{\theta_{\text{tp},i,j}\times\theta_{\text{fp}}\}]\\
&v_{i,j} \sim \text{Categorical}(\Omega_{\mathbf{i,j}})\\
&y_{i,j} \sim \text{Bernoulli}(1-\Omega_{1,i,j}).
\end{aligned}
$$

Here, $I(j \geq x_i)$ is an indicator function denoting whether occasion j is equal or greater than the specific time of species arrival at site i. Again, the specification of a process model for x_i and the redefinition of $\boldsymbol{\theta}_{\mathbf{tp}}$ are the only differences between the model above and the model presented by Chambert et al. (2015).

Only a small sample of the possible extensions are described above. Appendix S1 contains other examples and describes how to account for false positives across model types when using different approaches to estimate false positives.

Exploring Model Sensitivity to False-positive Incidence and the Number of Verified Samples

We undertook a simulation study to evaluate the baseline sensitivity of the RN and PA models to different amounts of false-positive error and the performance of extensions using the observation-confirmation protocol to account for false positives. We describe the simulation settings and present results below. Throughout, we fixed the simulated sampling effort as 200 sites with 20 sampling occasions each.

RN model

We first considered six different simulation scenarios representing combinations of across two abundance levels and three different incidences of false-positive error. We generated 300 replicate data sets per scenario with site-specific abundances $N_{i,\text{sim}} \sim \text{Poisson}(\lambda_{i,\text{sim}})$ and $\log(\lambda_{i,\text{sim}}) = \beta_0 + \beta_1 X_{1,i,\text{sim}}$, where $X_{1,i,\text{sim}} \sim N(0, 1)$, $\beta_0 = 0$ or -1.5 (three scenarios each), and $\beta_1 = 1$; sim indexes a particular simulation replicate. Thus, at a site with an average simulated covariate ($X_{1,i} = 0$), expected abundance was respectively roughly 0.23 animals or 1 animal. These values were chosen because the RN model tends to perform best when site-specific abundance is low (Kéry and Royle 2016:302) and it is perhaps most commonly applied to low-density species. We first generated "true" detection data as Bernoulli ($p_{i,\text{sim}}$), where $p_{i,\text{sim}} = 1-(1-r_{i,\text{sim}})^{N_{i,\text{sim}}}$, $\text{logit}(r_{i,\text{sim}}) = \alpha_0 + \alpha_1 X_{2,i,\text{sim}}$, $X_{2,i,\text{sim}} \sim N(0, 1)$, $\alpha_0 = -1.73$, and $\alpha_1 = 1$. Thus, an individual at an average site was expected to be detected with a probability of about 0.15 per sampling occasion.

Within these scenarios (Appendix S2: Table S1), we generated false-positive detections as occurring at random across all site intervals within a simulation (i.e., inclusive false positives). The probability of a false-positive detection within a cell was derived such that out of all detections, approximately 1%, 5%, or 10% were false positives with random Binomial sampling variance (absolute values of $\theta_{\text{fp,sim}}$ ranged from <0.001 to roughly 0.025). We defined θ_{fp} proportionally here rather than explicitly exploring specific values for the parameter itself because few studies provide empirical estimates of unconditional false-positive probabilities, many report percentages of observations that are true or false positives (e.g., Simons et al. 2007, Norouzzadeh et al. 2018), and proportional definitions are also commonly used to define thresholds for "accurate" data (e.g., 95% accuracy; Swanson et al. 2016).

Within each scenario, we sampled cells at random to create the verified data $v_{i,j,\text{sim}}$. We considered 10 levels for the number of site × occasion cells in which all detections were verified = {10, 20, 30, 40, 50, 60, 70, 80, 90 100}, with each scenario including 30 simulated data sets for each verification level. Each of the 1,800 generated data sets (300 replicate data sets for each of the six

simulation scenarios) was used to fit both a standard RN model and an extension accounting for false-positive error. For each estimator, we evaluated performance using mean error (absolute bias), root-squared error, standard deviation of the posterior distribution, coefficient of variation, and frequentist coverage (percentage of 95% credible intervals [CRIs] that included the true value) for **β, α**, and the finite-sample population size ($\hat{N}^{\text{tot}}$, derived as $\sum_{i=1}^{R}\hat{N}_i$; we considered relative [%] bias rather than mean error for this parameter). Because absence cannot be confirmed, allocating nondetections between **v** and **y** is an arbitrary decision: $\Pr(v_{i,j}=0)=\Pr(y_{i,j}=0)$, and we left all nondetections within **y** (Clare et al. 2019). We fit models here and below using JAGS v 4.0 (Plummer 2003) to perform Markov-chain Monte Carlo simulation through R v 3.4 (R Development Core Team 2017), although neither Bayesian estimation nor the complete data likelihood are prerequisite (Appendix S4 demonstrates fitting an extension of the RN model using maximum likelihood).

As expected, the RN model became increasingly biased as false positives constituted a greater proportion of detections (Fig. 2; Appendix S2). Random misclassification across all time periods and locations constituting 1%, 5%, or 10% of all detections led to respective relative biases of roughly 10%, 40%, and 70% across both simulated expected abundance levels. Models accounting for false-positive error exhibited less bias and root-mean-squared error (RMSE) regardless of the size of the verified sample. Estimator performance asymptotically improved as more samples were verified, with minimal improvement once detections within between 30 and 50 site × occasion cells were confirmed (Fig. 2; Appendix S2: Fig. S1).

PA model

Our exploration of the PA model was similar (Appendix S2: Table S5 outlines simulation scenarios). We first considered three scenarios with the following parameterization: $\text{logit}(\psi_{i,\text{sim}})=\beta_0+\beta_1 X_{1,i,\text{sim}}$, $X_{1,i,\text{sim}}\sim N(0,1)$, $\beta_0=0$, and $\beta_1=0.5$; $\text{logit}(p_{i,\text{sim}})=\alpha_0+\alpha_1 X_{2,i,\text{sim}}$, $X_{2,i,\text{sim}}\sim N(0,1)$, $\alpha_0=-2$, $\alpha_1=0.5$, and average arrival time φ = occasion 6. True observations $y_{i,j,\text{sim}}$ were generated as $\text{Bernoulli}(z_{i,\text{sim}}\times p_{i,\text{sim}}\times I(j\geq x_{i,\text{sim}}))$, where $z_{i,\text{sim}}\sim\text{Bernoulli}(\psi_{i,\text{sim}})$, and site- and simulation-specific arrival time $x_{i,\text{sim}}\sim\text{Poisson}(\varphi)$. That is, the occupancy probability at an average site was 0.50 and the probability of true detection conditional on arrival and occupancy at an average site was roughly 0.12. As before, we simulated 300 replicates per scenario, false-positive detections constituted 1%, 5%, or 10% of all detections, and the size of $v_{i,j,\text{sim}}$ ranged from 10 to 100. We fit both the standard PA model and the false-positive extension to each simulation replicate. We evaluated estimator properties with respect to **α**, **β**, $\hat{\varphi}$, and a finite-sample estimate of the proportion of occupied sites (PAO, derived for each simulation as $\sum_{i=1}^{R}\hat{z}_i$).

Results for the PA model largely mirrored those for the RN model. Across the range of false-positive detection proportions, bias in the estimated proportion of area occupied and arrival time when false-positive errors were ignored ranged from 3% to 20%, and 3% to 40%, respectively (Fig. 3, top panels). The extended model greatly reduced bias with any amount of verification effort, with little further reduction once 30 samples were confirmed (Fig. 3; Appendix S2: Fig. S2). Estimates of the proportion of area occupied were more uncertain when accounting for false-positive error, but more precise for arrival time when accounting for false-positive error (Fig. 3, bottom right), suggesting that false-positive error was inducing overdispersion in arrival estimates relative to Poisson expectations.

Evaluating the Suitability of Site-Confirmation Models

We briefly explored the consequences of lacking the ability to confirm false positives, and incorrectly assuming all false positives were conditionally exclusive (i.e., assuming false positives only occurred at unoccupied sites or sites prior to arrival, and ignoring false positives occurring at occupied sites or occurring concurrently with true positives). Following Chambert et al. (2015), we reconfigured the simulated data described above so that it mimicked the data collected under the site-confirmation design by considering only confirmed true positives. We fit the RN and PA models using site-confirmation approaches that incorrectly assumed false positives only occurred at unoccupied sites or occupied sites prior to arrival. These models were also generally unbiased given 30–50 confirmed true positives (Appendix S2: Figs. S3, S4, Tables S2, S6). However, the site-confirmation variant of the RN model occasionally (92 out of 1,800 simulated data sets) had difficulty converging, particularly when few true positives were confirmed and simulated abundance was small. The site-confirmation RN model also exhibited less than nominal frequentist coverage of finite-sample population size (Appendix S2: Tables S3, S4).

Model Sensitivity to Variability in False-positive Error

False-positive detections probably rarely happen randomly across space or time because the misclassified entities (other species or extrinsic phenomena) are not typically randomly distributed. Misclassification in detection/nondetection data is essentially spatiotemporal error in observed species occurrence, and the effects of false-positive error upon models using detection/nondetection data are analogous to the effects of spatial error upon models using presence/background or used/available data (Johnson and Gillingham 2008). If false positives always occur within the same sites and occasions as true positives, inference regarding the true positive

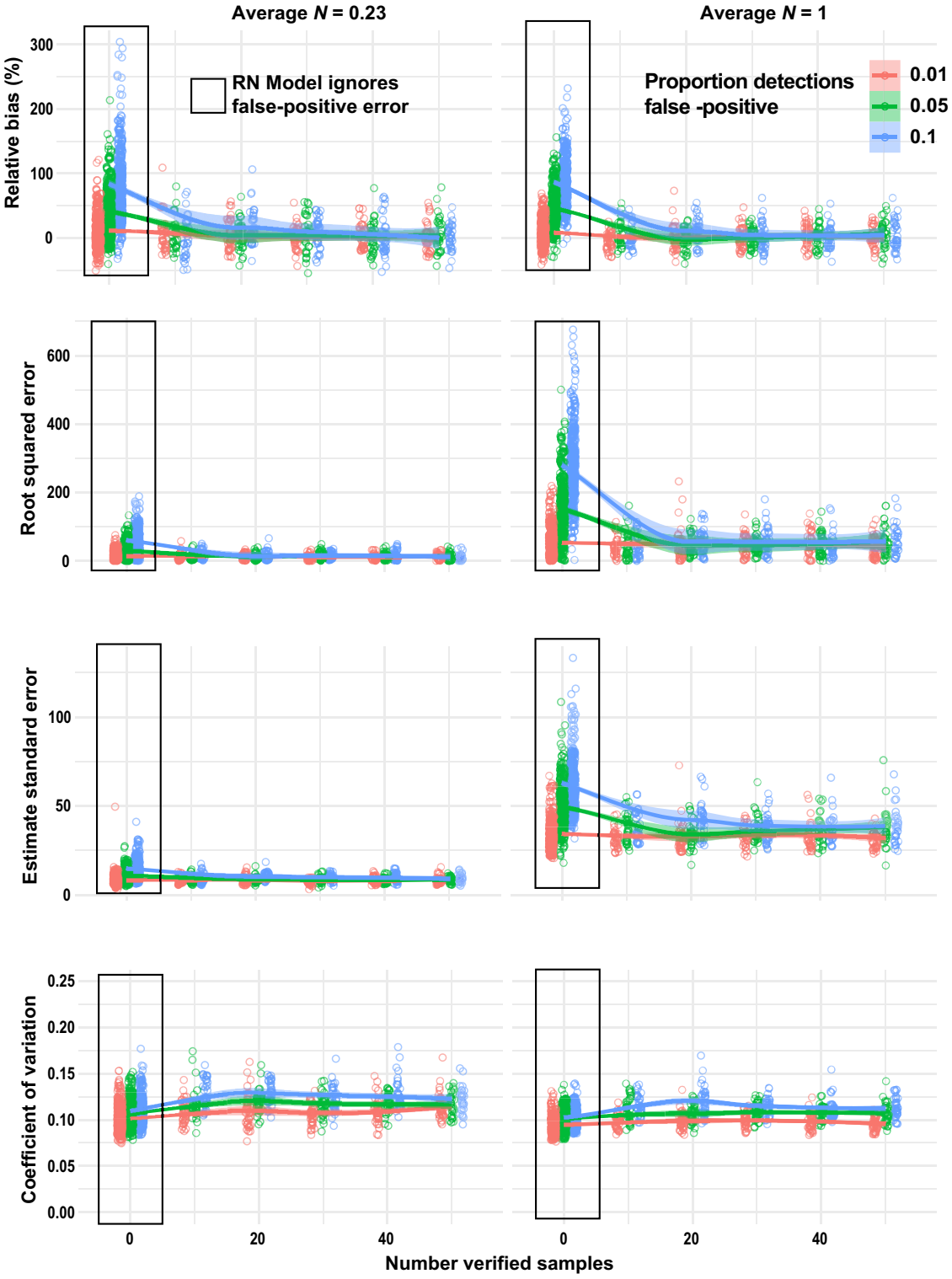

FIG. 2. Performance of the Royle–Nichols (RN) model ignoring false-positive error and of the model extension for false-positive error with regard to finite-sample population size under varying levels of random false-positive error (percentage of total detections = 1, 5, or 10) and verification effort (the number of sampling occasions in which all observations were verified). Standard error = standard deviation of the posterior distribution. The number of verified samples is truncated at 50 for visualization purposes (but see Appendix S2: Fig. S1). Smoothers depict means across different verification levels.

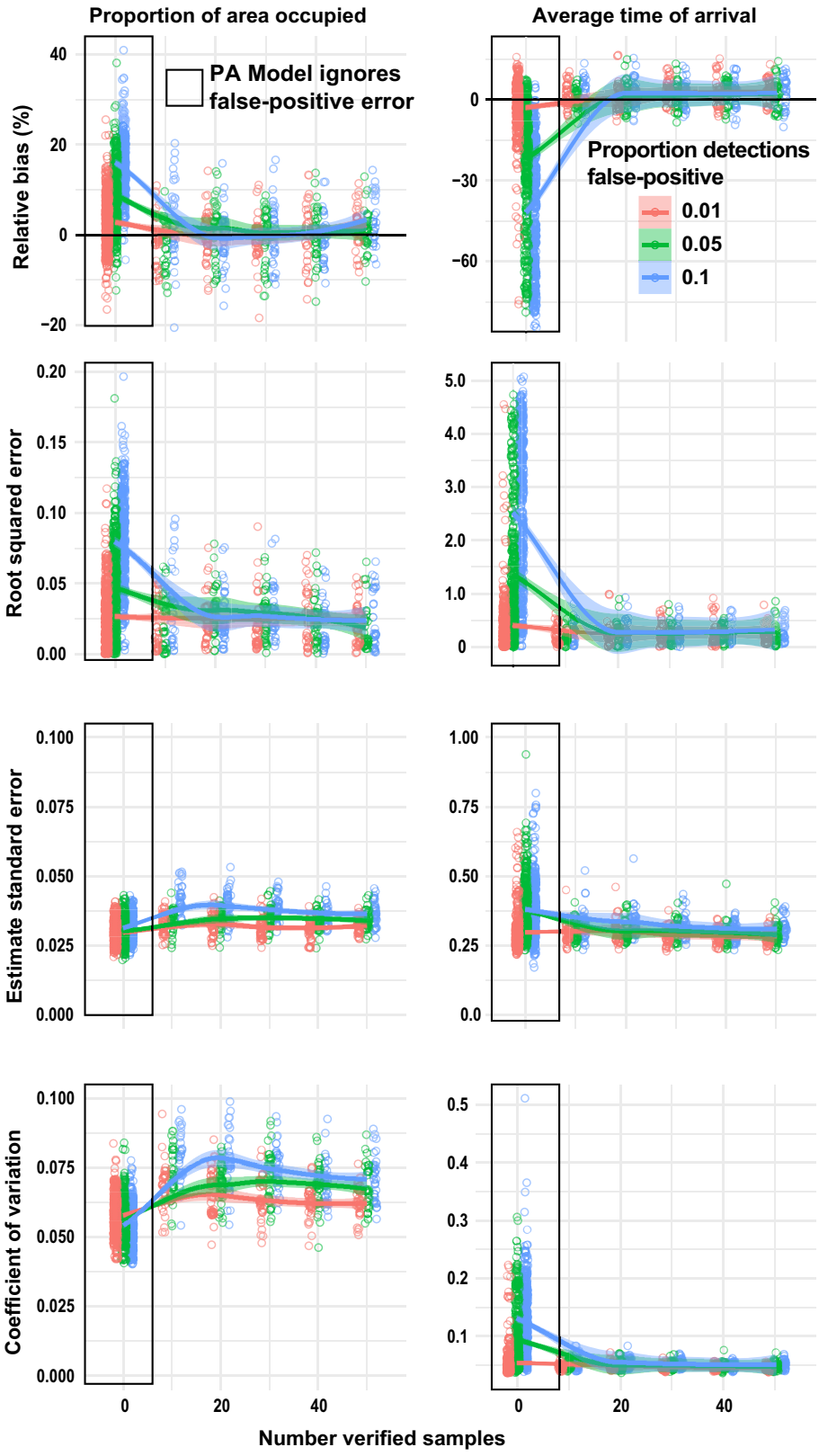

FIG. 3. Performance of a standard phenological occupancy model ignoring false-positive error and the model extension for

FIG. 3. (Continued)

false-positive error with regard to proportion of area occupied and the time of arrival under varying levels of random false-positive error (% of total detections = 1, 5, or 10) and verification effort (the number of sampling occasions in which all observations were verified). Standard error = standard deviation of the posterior distribution. The number of verified samples is truncated at 50 for visualization purposes (but see Appendix S2: Fig. S2). Smoothers depict means across different verification levels.

process is not affected. In general, one would expect false-positive error to result in larger bias if there were greater spatial or temporal (and by extension, environmental) separation between true- and false-positive detections.

RN model

We considered two subsequent scenarios (Appendix S2: Table S1) where $\beta_0 = -1.5$ and $\beta_1 = 1$ with $\boldsymbol{\alpha}$ defined as before, and $\text{logit}(\theta_{\text{fp},i}) = -6 + \{-1, 1\}X_{1,i,\text{sim}}$. As previously, $\log(\lambda_{i,\text{sim}}) = \beta_0 + \beta_1 X_{1,i,\text{sim}}$. Thus, false-positive observations were either more or less likely in locations with greater expected abundance and probability of true-positive observations. The verification protocol was simulated as previously. Empirically, simulated false-positive observations constituted about 6% of all observations. We fit three models to each scenario: one that assumed $\theta_{\text{fp}} = 0$, one that (incorrectly) assumed θ_{fp} was a constant in order to evaluate the consequences of ignoring variation in false-positive error, and one that (correctly) modeled $\theta_{\text{fp},i}$ as varying in relation to $X_{1,i,\text{sim}}$.

Abundance estimates using the standard RN model were more biased when false-positive detections were more likely to occur at locations where the focal species was less abundant and where true-positive detections were less likely (Appendix S2: Fig. S5). Abundance estimates produced by extensions accounting for false-positive error were nearly unbiased (<5% relative bias) regardless of the generating model for false-positive error, whether it was correctly specified within the fitted model, or the number of confirmed samples (Appendix S2: Fig. S5). Again, the performance of models accounting for false-positive error asymptotically improved when more samples were confirmed. Although models with a misspecified model for false-positive error improved more slowly, all models were essentially unbiased once 100 samples were confirmed.

PA model

We varied when and where false-positive errors occurred relative to the baseline settings within six further scenarios (Appendix S2: Table S5). In the first four scenarios, we defined $\theta_{\text{fp},i,j,\text{sim}} = \text{logit}^{-1}(-6 + \{-1, 1\}X_{1,i,\text{sim}})$ for $j > 4$ and $\theta_{\text{fp},i,j,\text{sim}} = 0$ for $j \leq 4$, and $\alpha_0 = \{-2, -1\}$. In the second two, $\theta_{\text{fp},i,j,\text{sim}} = \text{logit}^{-1}(-6 + \{-1, 1\}X_{1,i,\text{sim}})$ for $j > 2$ and $\theta_{\text{fp},i,j,\text{sim}} = 0$ for $j \leq 2$, and $\alpha_0 = -2$. Other values followed the previous description. That is, we used a shared covariate to make false positives either more or less likely to occur at occupied sites, and altered the timing of false positives so that rather than happening at any time, they initiated either one or three occasions before the average time of arrival. Empirically, these different formulations for false-positive error resulted in false positives accounting for between 3% ($\alpha_0 = -2$ and $\theta_{\text{fp},i,j,\text{sim}} = 0$ for $j \leq 4$) and 7% ($\alpha_0 = -1$ and $\theta_{\text{fp},i,j,\text{sim}} = 0$ for $j \leq 2$) of all detections. As with the RN model, we fit models that assumed $\theta_{\text{fp}} = 0$, models that (incorrectly) assumed θ_{fp} was a constant, and models that (correctly) assumed θ_{fp} varied in relation to $X_{1,i,\text{sim}}$.

The performance of the PA models accounting for false-positive error largely mirrored results seen with the RN model: there was little difference in model performance regardless of whether the probability of false-positive error was misspecified as a constant or allowed to vary spatially, and any differences shrank as the number of verified samples increased (Appendix S2: Figs. S6, S7). For models ignoring false-positive error, estimates of the proportion of area occupied became more biased as simulated false-positive error started earlier and when false positives were more likely in places where true positives were less likely (Appendix S2: Fig. S6). Moreover, occupancy estimation became more biased when the conditional probability of truly detecting a species was lower, indicating some potential sensitivity to small-sample bias.

When false positives were ignored, estimates of arrival time were insensitive to spatial patterns in error, but estimator bias increased as the simulated initiation of false positives occurred earlier relative to the average true arrival time (Appendix S2: Fig. S7). In fact, when false positives were simulated as starting only one sampling occasion before true positives, the estimator ignoring false positives exhibited less bias and smaller RMSE with respect to arrival time than the generating estimator. We believe this to be a specific case of offsetting biases induced by false-negative and false-positive error (see discussion in Appendix S2).

APPLICATION: PREDICTING GRAY FOX RELATIVE ABUNDANCE ACROSS WISCONSIN

Models for detection/nondetection data are often used to predict state variables spatially in order to prioritize management or conservation actions (Guélat and Kéry 2018). As a case study, we focus upon the relative abundance of gray fox (*Urocyon cinerargenteus*) in Wisconsin, USA, where its distribution is poorly understood. We used data from a monitoring program where trail-camera images are imperfectly classified via a crowdsourcing

platform (Clare et al. 2019) to investigate spatial patterns in fox relative abundance using the RN model. We modeled variation in fox expected abundance using a model accounting for false-positive error and one ignoring it. We used indicator variable selection (Kuo and Mallick 1998) to identify important predictors and regularize the log-linear coefficients within each model, and made statewide predictions for each by applying the model-averaged posterior predictive distribution across a 2 × 2 km lattice (more detail in Appendix S3).

Out of the images we reviewed, 67% were correctly classified; after further aggregation within 179 distinct 24-h sampling occasions (i.e., all detections within 179 occasions were reviewed), 60% consisted of only true positives, and 40% consisted of only false positives (either coyote, *Canis latrans* or red fox, *Vulpes vulpes*). Indicator variable selection provided less support for the inclusion of abundance covariates within the standard model than the model accounting for false-positive error, and the latter model suggested that false-positive error varied spatially in relation to the prevalence of surrounding cropland (Appendix S3: Tables S1, S2). Consequently, predictions from the standard model exhibit different spatial patterning (Fig. 4; although the statistical correlation between pairwise pixel estimates was fairly strong; $r = 0.80$). Furthermore, the point estimate for expected statewide population size derived via summation across the cells used for prediction was >3 times larger when false positives were ignored, although estimates overlapped substantially because of imprecision induced by the sparsity of observations. Although the estimated probability of a false-positive detection per sampling interval was very small (at an average site, 0.0015, 95% Credible Interval = 0.0012–0.0018; Appendix S3), this estimate suggests there were >100 false-positive detections across 91,276 total sampling occasions within the data set.

Discussion

Our results reiterate that when unaccounted for, false-positive detections can compromise a broad range of ecological applications and inferences. As expected, simulation results demonstrate that estimates of abundance and phenology can be biased by even moderate amounts of false positives. Although not tested via simulation, our case study suggests that biased parameter estimation can further lead to skewed spatial predictions. Estimation of abundance, which lacks a natural limit, appears particularly sensitive to false-positive error. When false positives were randomly generated, inclusive, and constituted 10% of all detections, the RN model (and a related unmarked spatial capture–recapture model, Appendix S1) exhibited 70% relative bias. For comparison, we have observed that occupancy models achieve this level of bias only when false-positive detections generated exactly in the same manner constitute 30% of all detections (Clare et al. 2019), and equal incidence of observationally or conditionally exclusive false-positive detections would be expected to induce yet greater bias (Miller et al. 2011). Given that detection/nondetection data often form the backbone of efforts to assess and

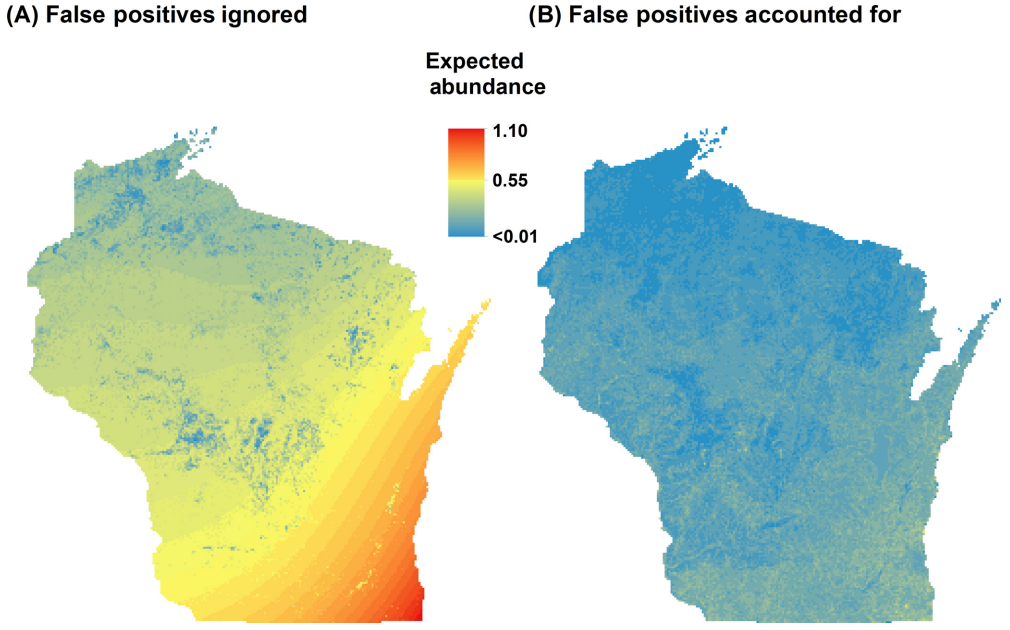

Fig. 4. Predictions of gray fox abundance across Wisconsin, USA in 2017 using a Royle–Nichols model assuming no false-positive error (left) and using an extension accounting for false-positive error (right).

monitor species populations or phenologies across large scales (Robinson et al. 2018, Jetz et al. 2019, Sun et al. 2019), techniques to ameliorate these errors are an important need. Luckily, our results suggest that existing model-based solutions designed for occupancy estimation are effective across a broad range of model classes that use detection/nondetection data.

We focused on model-based approaches following the observation-confirmation design. Researchers with the capability to confirm some subset of observations as true and false positives are broadly equipped to deal with false-positive error across different model classes by respecifying $\boldsymbol{\theta}_{\mathbf{tp}}$ to reflect the unconditional probability of detection within the model class of interest, specifying a process model for θ_{fp}, and collecting the necessary auxiliary data. We also demonstrated the extensibility of the site-confirmation design, and believe the calibration design is similarly flexible (Appendix S1). Two primary factors underlying the efficacy of these solutions relate to the number of confirmed or calibrated detections and how correctly the models for true and false positives are specified. We discuss these factors below, but preface by acknowledging that the extensibility of model-based approaches makes it challenging to quantify broadly the requisite confirmation effort or model structure for all possible applications. We encourage further simulation across a broader set of model classes (e.g., involving dynamics, data integration, or disease infection intensity; Miller et al. 2012*b*, Chandler and Clark 2014, Rossman et al. 2016) and study designs to help clarify these considerations.

The amount of auxiliary data required for unbiased estimation of the state variables of interest likely depends upon several factors. Our simulation results suggest that when false-positive error occurs at random and constitutes 10% or less of all detections, unbiased estimation following the site or observation-confirmation approaches may only require confirmed detections within 30–50 site-by-occasion intervals. Brost et al. (2018) demonstrate that the calibration design can be similarly reliable given 50 trials with known negatives. As the incidence of false-positive error is often less than 10% (McClintock et al. 2010), many applications may not require an exhaustive confirmation sample. More detections may need to be confirmed or calibrated to achieve unbiasedness if false positives are more common or the observed data is sparse (Ruiz-Gutiérrez et al. 2016). Because sampling efficiency and the incidence of false-positive error are often difficult to gauge a priori, investigators may be better informed by confirming as many samples as feasible during initial project phases in order to buffer against uncertainty (Clement 2016). As such, perhaps the most important factor to consider when choosing a protocol to account for false-positive error is which approach is likely to generate the largest amount of auxiliary data with the least effort or expense (Chambert et al. 2015, Ruiz-Gutiérrez et al. 2016). Note that a single confirmation under the observation-confirmation protocol requires that all observations at a site and specific occasion have been confirmed as true, or that all have been confirmed as false positives, or that >0 true- and false-positive observations have been confirmed. A confirmation following the site-confirmation protocol is simply any site by occasion in which >0 true positives have been confirmed: these data may be substantially easier to generate in certain sampling situations.

A connected reason to assimilate more auxiliary data is to improve modeling variation in where and when false positives occur. Modeling variation in $\boldsymbol{\theta}_{\mathbf{fp}}$ using covariates or spatiotemporal dependence terms can account for potential estimator biases associated with missing heterogeneity (Miller et al. 2015), and can be critical for reliably predicting spatial patterns or trends in species distributions (sensu Guélat and Kéry 2018, case study here). Furthermore, as our simulations demonstrate, understanding the covariance between true- and false-positive detections can provide insights into the amount of estimator bias likely to be induced by false positives. Although the observation-confirmation design appeared robust to misspecifying the false-positive process when confirmation followed a random sample and the true-positive process was correctly modeled, it seems unlikely that all applications will be as robust. Appropriately modeling variation in true- and false-positive processes may be particularly critical if the data at hand provide little capacity to differentiate true and false positives. For example, fully latent estimators are particularly sensitive to model structure (Miller et al. 2015). We note these performed well when applied to RN and PA models when provided informed priors for false-positive parameters and when all generating processes were properly parameterized (Appendix S2). However, although we agree with Miller et al. (2015) that such approaches deserve more consideration if no other options are available, we strongly recommend confirmation or calibration if possible to buffer against the risk of choosing poor prior distributions or specifying poor models for false-positive error.

Assumptions regarding where and when false positives can occur may also deserve further consideration. Because the false-positive process can be directly quantified, the observation-confirmation and calibration designs allow investigators to assume true- and false-positive observations are either inclusive or conditionally exclusive. The site-confirmation design appears to require further constraints to address observation-level false positives (Appendix S1). Assuming that false positives can only occur in situations where true positives are impossible constrains the range of possible observation outcomes, adding precision and making certain approaches identifiable. However, violations of the assumption may carry costs. Here, the PA model was unaffected, but uncertainty intervals for finite-sample population size using the site-confirmation RN model were permissive. Brost et al. (2018) found that

occupancy estimation remained nearly unbiased but exhibited poor coverage because estimates of true-positive detection given occurrence were positively biased. There are likely to be situations where such assumption violations bias ecological parameters as well as observational parameters. For example, if using the RN model to infer abundance at a set of locations that all happened to be occupied, assuming only site-level false positives would not be effective.

The models described here are designed to account for varied types of false-positive error across a range of sampling techniques (Chambert et al. 2015). A cost of this generality is that they are not tuned for specific sampling problems. For example, they do not distinguish between different types of misclassification—for example, pairwise misclassification of different species or individuals would be useful within multispecies or capture–recapture models—and do not leverage other specific information such the count of observations within an occasion (e.g., Conn et al. 2013, Chambert et al. 2018, Augustine et al. 2020). Where appropriate, these fine-tuned solutions may be preferable, and can often also be extended to other estimation problems by refining the state process.

Our motivation for pursuing generalizable model-based solutions was grounded in concerns regarding the efficiency of implementing alternative solutions within our own work. Simulation results here reinforce our concerns. Bias associated with false-positive error depends on the model employed, the incidence of error, and where and when false positives occur. As such, it may only be safe to ignore false-positive detections when making ecological inference if they have extremely low incidence or generally happen at the same time and place as true positive detections. These conditions are difficult to ascertain without collecting information about classification performance that itself could be used to develop a model-based solution (Ruiz-Gutiérrez et al. 2016, Clare et al. 2019). The ability to leverage the efficiency of model-based solutions across a broader range of model classes should make it substantially easier for investigators to account for the false-positive errors that pervade ecological data.

Acknowledgments

Support for this research was provided by NASA ESSF NNX16AO61H and Ecological Forecasting grant NNX14AC36G, and a grant from the Federal Aid in Wildlife Restoration act awarded to Wisconsin Department of Natural Resources. We use data partially generated via the Zooniverse.org platform funded by a grant from the Alfred P. Sloan Foundation and a Global Impact Award from Google. Comments from V. Ruiz-Gutiérrez and four anonymous reviewers improved the manuscript.

Literature Cited

Augustine, B. C., J. A. Royle, D. W. Linden, and A. K. Fuller. 2020. Spatial proximity moderates genotype uncertainty in genetic tagging studies. Proceedings of the National Academy of Sciences USA 117:17903–17912.

Brost, B. M., B. A. Mosher, and K. A. Davenport. 2018. A model-based solution for observational errors in laboratory studies. Molecular Ecology Resources 18:580–589.

Chambert, T., D. A. W. Miller, and J. D. Nichols. 2015. Modeling false positive detections in species occurrence data under different study designs. Ecology 96:332–339.

Chambert, T., J. H. Waddle, D. A. W. Miller, S. C. Walls, and J. D. Nichols. 2018. A new framework for analyzing acoustic species detection data: occupancy estimation and optimization of recordings post-processing. Methods in Ecology and Evolution 9:560–570.

Chandler, R. B., and J. D. Clark. 2014. Spatially-explicit integrated population models. Methods in Ecology and Evolution 5:1351–1360.

Clare, J. D. J., P. A. Townsend, C. Anhalt-Depies, C. Locke, J. L. Stenglein, S. Frett, K. J. Martin, A. Singh, T. R. Van Deelen, and B. Zuckerberg. 2019. Making inference with messy (citizen science) data: when are data accurate enough and how can they be improved? Ecological Applications 29: e01849.

Clement, M. J. 2016. Designing occupancy studies when false-positive detections occur. Methods in Ecology and Evolution 7:1538–1547.

Conn, P. B., B. T. McClintock, M. F. Cameron, D. S. Johnson, E. E. Moreland, and P. L. Boveng. 2013. Accommodating species identification errors in transect surveys. Ecology 94:2607–2618.

Ferguson, P. F. B., M. J. Conroy, and J. Heppinstall-Cymerman. 2015. Occupancy models for data with false positive and false negative errors and heterogeneity across sites and surveys. Methods in Ecology and Evolution 6:1395–1406.

Gardiner, M. M., L. L. Allee, P. M. Brown, J. E. Losey, H. E. Roy, and R. R. Smyth. 2012. Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. Frontiers in Ecology and the Environment 10:471–476.

Guélat, J., and M. Kéry. 2018. Effects of spatial autocorrelation and imperfect detection on species distribution models. Methods in Ecology and Evolution 9:1614–1625.

Jetz, W., et al. 2019. Essential biodiversity variables for mapping and monitoring species distributions. Nature Ecology and Evolution 3:539–551.

Johnson, C. J., and M. P. Gillingham. 2008. Sensitivity of species distribution models to error, bias, and model design: an application to resource selection functions for woodland caribou. Ecological Modeling 213:143–155.

Kéry, M., and J. A. Royle. 2016. Applied hierarchical modeling in ecology. Volume 1. Academic Press, London, UK.

Kosmala, M., A. Wiggins, A. Swanson, and B. Simmons. 2016. Assessing data quality in citizen science. Frontiers in Ecology and the Environment 14:551–560.

Kuo, L., and B. Mallick. 1998. Variable selection for regression models. Sankhyā 60:65–81.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248–2255.

McClintock, B. T., L. L. Bailey, K. H. Pollock, and T. R. Simons. 2010. Experimental investigation of observation error in anuran call surveys. Journal of Wildlife Management 74:1882–1893.

Miller, D. A. W., L. L. Bailey, E. H. C. Grant, B. T. McClintock, L. A. Weir, and T. R. Simons. 2015. Performance of species occurrence estimators when basic assumptions are not met: a

test using field data where true occupancy status is known. Methods in Ecology and Evolution 6:557–565.

Miller, D. A., J. D. Nichols, B. T. McClintock, E. H. Campbell Grant, L. L. Bailey, and L. A. Weir. 2011. Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. Ecology 92:1422–1428.

Miller, D. A. W., L. A. Weir, B. T. McClintock, E. H. Campbell Grant, L. L. Bailey, and T. R. Simons. 2012*a*. Experimental investigation of false positive errors in auditory species occurrence surveys. Ecological Applications 22:1665–1674.

Miller, D. A. W., B. L. Talley, K. R. Lips, and E. H. Campbell Grant. 2012*b*. Estimating patterns and drivers of infection prevalence and intensity when detection is imperfect and sampling error occurs. Methods in Ecology and Evolution 3:850–859.

Norouzzadeh, M. S., A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune. 2018. Automatically identifying, counting, and describing wildlife animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences 115:E5716–E5725.

Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using GIBBS sampling. Pages 20–22 *in* Proceedings of the 3rd International Workshop on Distributed Statistical Computing. http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Drafts/Plummer.pdf

R Development Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.r-project.org

Ramsey, D. S. L., P. A. Caley, and A. Robley. 2015. Estimating population density from presence–absence data using a spatially explicit model. Journal of Wildlife Management 79:491–499.

Robinson, O. J., V. Ruiz-Gutiérrez, D. Fink, R. J. Meese, M. Holyoak, and E. G. Cooch. 2018. Using citizen science data in integrated population models to inform conservation. Biological Conservation 227:361–368.

Rossman, S., C. B. Yackulic, S. P. Saunders, J. Reid, R. Davis, and E. F. Zipkin. 2016. Dynamic N-occupancy models: estimating demographic rates and local abundance from detection/non-detection data. Ecology 97:3300–3307.

Roth, T., N. Strebel, and V. Amrhein. 2014. Estimating unbiased phenological trends by adapting site-occupancy models. Ecology 95:2144–2154.

Royle, J. A., and W. A. Link. 2006. Generalized site occupancy models allowing for false positive and false negative errors. Ecology 87:835–841.

Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence–absence data or point counts. Ecology 84:777–790.

Ruiz-Gutiérrez, V., M. B. Hooten, and E. H. Campbell Grant. 2016. Uncertainty in biological monitoring: a framework for data collection and analysis to account for multiple sources of sampling bias. Methods in Ecology and Evolution 7:900–909.

Simons, T. R., M. W. Alldredge, K. H. Pollock, and J. M. Wettroth. 2007. Experimental analysis of the auditory detection process on avian point counts. Auk 124:986–999.

Sun, C. C., J. A. Royle, and A. K. Fuller. 2019. Incorporating citizen science data in spatially explicit integrated population models. Ecology 100:e02777.

Swanson, A., M. Kosmala, C. Lintott, and C. Packer. 2016. A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. Conservation Biology 30:520–531.

Supporting Information

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/ecy.3241/suppinfo

Data Availability

Data and code are archived with Zenodo: http://doi.org/10.5281/zenodo.3862705.